

# VU Research Portal

## Evolution in games with a continuous action space

van Veelen, C.M.

2001

### **document version**

Early version, also known as pre-print

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

van Veelen, C. M. (2001). *Evolution in games with a continuous action space*. (TI Discussion Paper; No. 2001/068/1). Econometrie (ECTRIE).

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)



TI 2001-068/1

Tinbergen Institute Discussion Paper

# Evolution in Games with a Continuous Action Space

*Matthijs van Veelen*

*Econometrics and Operational Research, Faculty of Economics and Business Administration, Vrije  
Universiteit Amsterdam, and Tinbergen Institute*

**Tinbergen Institute**

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

**Tinbergen Institute Amsterdam**

Keizersgracht 482  
1017 EG Amsterdam  
The Netherlands  
Tel.: +31.(0)20.5513500  
Fax: +31.(0)20.5513555

**Tinbergen Institute Rotterdam**

Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31.(0)10.4088900  
Fax: +31.(0)10.4089031

Most TI discussion papers can be downloaded at  
<http://www.tinbergen.nl>

# Evolution in games with a continuous action space.

Matthijs van Veelen  
Tinbergen Instituut  
Keizersgracht 482  
1017 EG Amsterdam

July 20, 2001

## Abstract

Allowing for games with a continuous action space, we deal with the question whether and when static concepts like evolutionary stability can shed any light on what happens in the dynamical context of a population playing these games. The continuous equivalents of theorems for the finite case are either harder to prove or simply untrue. In some cases that fall within the latter category, sensible additional assumptions can repair the damage. Apart from that it turns out that we can no longer ignore the mutation process; with a continuous action space it makes quite a difference what kind of shocks we consider to be likely.

## 1 Introduction

Evolutionary game theory mainly focusses on symmetric games with a discrete and even finite action space. For most of the phenomena we seek to explain that is no real restriction; in quite some cases these games fit reality well enough to help us say something sensible about human or animal behaviour. For the finite case a coherent theory has been developed, which is well reflected in the book by Weibull (1995). There may be evolutionary interesting games though that naturally have a continuous strategy space. Trying to tackle these games as well, useful results have been proven by, amongst others, Seymour (2000) and Oechssler and Riedel (2000, 2001), but the theory is still not as complete as for the finite case. This paper is an attempt to get a bit closer to a complete picture by gathering the more significant results and, more than that, finding the missing links.

Like Seymour and Oechssler and Riedel, we will closely follow the concepts of evolutionary stability and replicator dynamics as they are defined for the finite case and our aim is to find out whether these two can be linked. We will discover that continuity complicates matters; definitions that coincide in the discrete case

cease to do so in the continuous case. As a result of this evolutionary stability in its most common definition no longer guarantees asymptotic stability in the replicator dynamics. The main part of this paper - section 3 - is therefore devoted to sorting out which of the standard results hold for the continuous case as they did in the finite and which do not. Some of the results that do not survive the transition can be restored under additional assumptions.

Another complication is that with a continuous action space it is a lot less clear what *small* perturbations of a population state are; there is a variety of possibilities what kind of shocks we think a population should be able to overcome before we call it stable. Section 4 lists the different options and discusses how well they fit a model of mutation. Section 5 is devoted to an alternative approach, taken by Sigmund and Hofbauer (1998). Their adaptive dynamics may for a subset of games prescribe dynamical behaviour similar to the behaviour of our replicator dynamics, but there are differences that need to be brought out.

## 2 Definitions

Before we start I would like to stress the fact that we do allow for mixed strategies here. This possibility is sometimes ignored in games with a continuous action space for obvious reasons: no mixed strategies are needed to prove existence of an equilibrium of a game with a compact and convex action space and continuous and quasiconcave utility functions. But just the fact that we do not need the possibility of mixed strategies to make sure an equilibrium exists does not make mixed strategies any more or less realistic. A mixed strategy can be seen as either to represent a player that draws his action from a distribution or as a population in which different actions are being played and in which one is uncertain which player one will encounter. The last interpretation is obviously the one evolutionary game theory favours. A (symmetric) equilibrium in mixed strategies is then nothing but a heterogeneous population being in equilibrium with itself, and I think there is no reason why we should not consider this possibility in the continuous case.

Let's start with the game. The action space  $S$  is a subset of  $\mathbb{R}$  and at some points in the text  $S$  will be assumed to be compact and/or convex. A strategy is being described with a probability measure  $P$  on  $(S, \mathcal{B})$  with Borel  $\sigma$ -field  $\mathcal{B}$ . We assume all probability measures not to be singular and the set of all those non-singular probability measures on  $S$  is denoted by  $\mathcal{P}[S, \mathcal{B}]$ . Sometimes there is reason to restrict ourselves to distributions that have a density, that is, probability measures  $P$  for which there is a function  $f : S \rightarrow \mathbb{R}$  such that  $P(B) = P\{x \in B\} = \int_B f(y) dy$  for any  $B \in \mathcal{B}$ , but in general we do allow for mixtures. Therefore, for any  $P$  we will define  $\mathcal{A}_P$  as its set of atoms and  $\mathcal{C}_P$  as the support set of the remainder. To be a bit more formal, we first define  $\mathcal{A}_P = \{x \in S | P(\{x\}) > 0\}$ . Since we only consider non-singular probabilities, we know that there is a function  $f : S \rightarrow \mathbb{R}$  with  $\int_S f(y) dy = 1 - P(\mathcal{A}_P)$  such that  $P(B \setminus \mathcal{A}_P) = \int_B f(y) dy$  and with this function we define  $\mathcal{C}_P = \text{Cl}\{x \in S | f(x) > 0\}$ . Obviously we can equate the support

set  $C(P)$  of  $P$  with  $\mathcal{A}_P \cup \mathcal{C}_P$ .

The situation we will consider mostly is one of pairwise contests and random matching; players are drawn at random from a population to play a symmetric 2-player game, which is characterized by a function  $A : S \times S \rightarrow \mathbb{R}$ . Here  $A(x, y)$  is the payoff to player 1 if he plays  $x$  and the opponent plays  $y$ . Symmetry implies that the payoff to player 2 then is  $A(y, x)$ . The expected payoff to a player playing strategy  $P$  against a player playing  $Q$  is given by the function  $u[P, Q] = \int \int A(x, y) dP(x) dQ(y)$ . One could think of more general forms of this function  $u[P, Q]$  that are not necessarily bilinear. This would allow for games like the sex ratio game as described by Sigmund (1987), but it should be noted that any expected payoff function that is not bilinear implies a departure from the setting of pairwise contests, as Bomze and Pötscher (1989) show. Most of the results in this paper remain valid though in the more general setting where  $u[P, Q] = \int F_Q(x) dP(x)$ .

The following three definitions are close copies of their discrete counterparts.

**Definition 1**  *$P$  is an evolutionary stable strategy if there exists an invasion barrier  $\epsilon_Q$  for every strategy  $Q \neq P$ ;  $\epsilon_Q$  is an invasion barrier for  $Q \neq P$  if  $u[P, (1 - \epsilon)P + \epsilon Q] > u[Q, (1 - \epsilon)P + \epsilon Q]$  for all  $\epsilon \in (0, \epsilon_Q)$*

An equivalent definition of evolutionary stability is that both  $u[P, P] \geq u[Q, P]$  for all  $Q$  and  $u[P, P] = u[Q, P] \Rightarrow u[P, Q] > u[Q, Q]$  for all  $Q \neq P$  must hold.

**Definition 2**  *$P$  has a uniform invasion barrier if there exists an  $\bar{\epsilon} \in (0, 1)$  such that  $u[P, (1 - \epsilon)P + \epsilon Q] > u[Q, (1 - \epsilon)P + \epsilon Q]$  for all  $Q \neq P$  and all  $\epsilon \in (0, \bar{\epsilon})$*

**Definition 3**  *$P$  is locally superior if it has a vicinity  $V$  such that  $u[P, Q] > u[Q, Q]$  for all  $Q \neq P$  in  $V$ .*

Whether or not a strategy is locally superior obviously depends on what we consider to be a vicinity. In this section I will give the bare definitions of a few types of vicinities, but the question what vicinities we allow for is not just of technical interest. After all, it would be nice if we could see a vicinity of a population state  $P$  as a set of states that can be reached from  $P$  by mutations with relatively large probability. Here we will stick to the technicalities, but Section 4 will be entirely devoted to the interpretation of the types of vicinities. Until then we will assume they are all equally sensible.

Please do note that we use the word *vicinity* instead of the term *neighbourhood*. A vicinity is defined with the help of a function  $d : \mathcal{P}[S, \mathcal{B}] \times \mathcal{P}[S, \mathcal{B}] \rightarrow \mathbb{R}_0^+$ , which is not necessarily a distance;  $V$  is a vicinity of  $P$  if and only if there is a  $\delta > 0$  such that  $V = \{Q \in \mathcal{P}[S, \mathcal{B}] \mid d(P, Q) < \delta\}$ . If this function  $d$  is a distance, then the set of all vicinities of all population states  $P$  makes a basis for a topology. However, one of the examples that follow is not a distance and therefore it is better to avoid using neighbourhoods, that are only defined once there is a topology. The details of the distinction between vicinities and

neighbourhoods can be found in the appendix, that also provides the proper topological context.

We consider three possible functions  $d$ , the first one being the Prohorov metric, which is defined by

$$d_{pr}(P, Q) = \inf \left\{ \varepsilon > 0 \mid \begin{array}{l} P(B) \leq Q(B^\varepsilon) + \varepsilon, \\ Q(B) \leq P(B^\varepsilon) + \varepsilon \end{array} \text{ for all } B \in \mathcal{B} \right\}$$

where  $B^\varepsilon = \{x \in S \mid \exists y \in B \text{ such that } |x - y| < \varepsilon\}$ .<sup>1</sup>

It is worth knowing that the Prohorov metric metrizes weak convergence, that is,  $d_{pr}(P_n, P) \rightarrow 0$  if and only if  $\int g(x) dP_n(x) \rightarrow \int g(x) dP(x)$  for all bounded and continuous  $g$ . The topology it induces is called the weak topology.

The second option is the variational distance,

$$d_{vd}(P, Q) = 2 \sup \{|P(A) - Q(A)| \mid A \in \mathcal{B}\}$$

$$= \int \left| \frac{dP}{d\mu} - \frac{dQ}{d\mu} \right| d\mu$$

for any  $\mu$  dominating both  $P$  and  $Q$ ,

which induces the strong topology. To illustrate the difference between the vicinities we make with those two functions, we can take  $S = [0, 1]$  and  $P$  degenerate in 0. Now compare the collection  $\mathcal{V}_{pr}(P)$  of all vicinities of  $P$  in the Prohorov metric to  $\mathcal{V}_{vd}(P)$  that contains all vicinities of  $P$  in the variational distance. It is not too hard to see that these two collections of vicinities share no elements apart from the whole set  $\mathcal{P}[S, \mathcal{B}]$ ; just take a look at the probability measures  $P_y$  degenerate in  $y$ . For  $y \in (0, 1]$  we find that  $d_{vd}(P, P_y) = 2$  whereas  $d_{pr}(P, P_y) = y$ . Therefore any vicinity of  $P$  in the Prohorov metric contains probability measures  $P_y$  (choose  $y < \delta$  for a vicinity  $V_{pr}(P, \delta) = \{Q \in \mathcal{P}[S, \mathcal{B}] \mid d_{pr}(P, Q) < \delta\}$ ) while none of these  $P_y$  is contained in any of the vicinities of  $P$  in the variational distance, apart from, of course, the whole set  $\mathcal{P}[S, \mathcal{B}]$  which is a vicinity in the variational distance for  $\delta > 2$  and in the Prohorov metric for  $\delta \geq 1$ .

Going in the opposite direction is more succesful; because  $d_{pr}(P, Q) \leq \frac{1}{2}d_{vd}(P, Q)$  we find that  $V_{vd}(P, 2\delta) \subset V_{pr}(P, \delta)$ .

As a third possibility I suggest to use the Kullback-Leibler relative-entropy measure:

$$d_{kl}(P, Q) = H_P(Q) = \int_{C(P)} \log \left( \frac{dP(x)}{dQ(x)} \right) dP(x)$$

---

<sup>1</sup> See for instance Bickel, Klaassen, Ritov and Wellner (1993)

This function does not satisfy the triangle inequality and is not symmetric. In the appendix we show that the set of all vicinities of all distributions  $P \in \mathcal{P}[S, \mathcal{B}]$  can not be a basis for a topology. In an alternative way one can use this measure to define a non-trivial topology that could go by the name 'very strong' topology. However, in this topology not all distributions have a neighbourhood other than the whole set  $\mathcal{P}[S, \mathcal{B}]$  and that is the reason why we prefer to use vicinities, of which there are enough for every distribution  $P \in \mathcal{P}[S, \mathcal{B}]$ . Again, for details I refer to the appendix.

To see where the vicinities we make with this Kullback-Leibler relative-entropy measure differ from the other two, take  $S = [0, 1]$  and  $P$  uniform on  $S$ . Again we make a collection  $\mathcal{V}_{kl}(P)$  of all vicinities of  $P$  in the Kullback-Leibler entropy and compare this to  $\mathcal{V}_{vd}(P)$ . If we then consider distributions  $P_y$  uniform on  $[y, 1]$ , we find that for all  $y \in (0, 1)$  the Kullback-Leibler entropy  $d_{kl}(P, P_y) = \infty$  while  $d_{vd}(P, P_y) = 2y$ . This implies that  $\mathcal{V}_{kl}(P)$  and  $\mathcal{V}_{vd}(P)$  have no elements in common; any vicinity of  $P$  in the variational distance contains probability measures  $P_y$  (just choose  $y < \delta/2$  for a vicinity  $V_{vd}(P, \delta) = \{Q \in \mathcal{P}[S, \mathcal{B}] \mid d_{vd}(P, Q) < \delta\}$ ) while none of these  $P_y$  is contained in any of the vicinities of  $P$  in the Kullback-Leibler entropy since  $d_{kl}(P, P_y) = \infty$ .

Bomze (1991) uses Reiss (1989) to show that  $[d_{vd}(P, Q)]^2 \leq d_{kl}(P, Q)$  and therefore we know that  $V_{kl}(P, \delta^2) \subset V_{vd}(P, \delta)$ .

Summing up the relation between these three types of vicinities in an intuitive way, one can say that  $Q$  being close to  $P$  in the variational distance implies that  $Q$  is close to  $P$  in the Prohorov metric, but not the other way round, and that if  $Q$  is close to  $P$  in the Kullback-Leibler sense,  $Q$  will also be close to  $P$  in the variational distance, but not the other way round.

In this paper I chose to stick to the term local superiority and mention what types of vicinities are used. In the literature, some of those kinds of local superiorities got names of their own; Bomze (1991) calls a strategy that is locally superior in the variational distance *strongly uninvadable* and Oechssler and Riedel (2000) term a strategy that is locally superior in the Prohorov metric *evolutionary robust*.

For the derivation of the *replicator dynamics* we can simply follow the standard case. A population at time  $t$  is characterized by a measure  $R(t)$ , where  $R(t)(B)$  is the magnitude of the part of the population playing a strategy in  $B \in \mathcal{B}$  at time  $t$ . By definition the expected payoff of a strategy in the game determines the expected growth of the part of the population that plays it, so we get

$$\frac{d}{dt}(R(t)(B)) = \int_B \int_S A(x, y) P(t)(dy) R(t)(dx) \quad \forall B \in \mathcal{B}$$

$$\text{where } P(t)(C) = \frac{R(t)(C)}{R(t)(S)} \quad \forall C \in \mathcal{B}$$

which can be rewritten as



$$\begin{aligned} \frac{d}{dt} (P(t)(B)) = \\ \int_B \int_S A(x, y) P(t)(dy) P(t)(dx) - P(t)(B) \int_S \int_S A(x, y) P(t)(dy) P(t)(dx) \end{aligned} \quad \forall B \in \mathcal{B}$$

In the more general, non-bilinear case this would be:

$$\frac{d}{dt} (P(t)(B)) = \int_B F_{P(t)}(x) P(t)(dx) - P(t)(B) \int_S F_{P(t)}(x) P(t)(dx) \quad \forall B \in \mathcal{B}$$

Note that the space on which the last two differential equations are defined is the set  $\mathcal{P}[S, \mathcal{B}]$  since  $P(t)(S) = 1$ . All points along a trajectory through a point  $P(0)$  have the same atoms, but *in the limit* atoms can both pop up and disappear. More formally, if we define  $\mathcal{A}_{P(\infty)} = \{x \in S \mid \lim_{t \rightarrow \infty} P(t)(\{x\}) > 0\}$  then neither  $\mathcal{A}_{P(0)} \subset \mathcal{A}_{P(\infty)}$  nor  $\mathcal{A}_{P(\infty)} \subset \mathcal{A}_{P(0)}$  necessarily holds.

The states that interest us most are the asymptotically stable states; states that guarantee a pull back to status quo after any small perturbation of the population state. Like local superiority, this concept is dependent on what we consider small perturbations, or in other words: which type of vicinities we use. As mentioned, Section 4 will be devoted to considerations of that kind. But apart from that, asymptotic stability can be an empty concept for some choices of vicinities. This will be discussed at the end of Section 3.

Of course there are other dynamics on a population we could look at; the standard replicator dynamics is one out of many payoff monotonic growth-rate functions. But given that we restrict ourselves to the setting of games being played within one population,<sup>2</sup> I think that the replicator dynamics is the only interesting dynamics. The reason for this is the following. Consider a function  $O : \mathcal{P}[S, \mathcal{B}] \rightarrow \mathcal{P}[S, \mathcal{B}]$  which defines a dynamical system

$$\frac{d}{dt} P = O(P)$$

Then either there is a game  $A : S \times S \rightarrow \mathbb{R}$  for which the replicator dynamics gives this  $O$  or there is not.<sup>3</sup> In case there is, this  $A$  is unique up to an affine transformation, and although the same dynamics  $O$  could be composed of another game  $A'$  that is not an affine transformation of  $A$  and another dynamics, the proper thing to do then would be to rescale  $A'$  until it does fit the replicator dynamics and not the other way round; the payoffs are to reflect fitness and

---

<sup>2</sup>One could indeed allow for players from one population playing games with players from another population. In that case there is the generalized replicator dynamics, that falls within the class of *positive definite adaptive* dynamics. (Seymour (2000)). This multi-population setting though is a setting we do not consider here.

<sup>3</sup>Again, for the case where we do not assume pairwise contest, replace the game  $A$  by a function  $F$  that assigns to every element of  $\mathcal{P}[S, \mathcal{B}]$  a function on the actionspace, that is,  $F_Q : S \rightarrow \mathbb{R}, Q \in \mathcal{P}[S, \mathcal{B}]$  and use the non-bilinear replicator dynamics.

Figure 1:

that is what they do if they combine with the replicator dynamics to the given dynamical system. If, on the other hand, there is no such  $A$  that does this trick, this setting could never do as an explanation, so that puts such a dynamical system outside the scope of this paper.

### 3 From evolutionary to asymptotically stable and back

The most interesting question is now whether we can expect static concepts like evolutionary stability to be of any use when trying to figure out where the harder to handle replicator dynamics would take us. We know that in the discrete case evolutionary stability (ESS), the existence of a uniform invasion barrier (UIB) and local superiority (LS) are equivalent and imply asymptotic stability (AS). Continuity breaks up a few of those links as we will see just now. The examples that show which implications do not hold are of course negative results that suggest that local superiority is the most useful of the three static concepts, but all is not lost for evolutionary stability; there are assumptions that pave the way from evolutionary stability to local superiority.

The implications under examination are drawn as arrows in figure 1. In this section we will check them one at a time trying to find out which hold and which don't. We start with the relation between evolutionary stability and the existence of a uniform invasion barrier.

#### 1 ESS $\nRightarrow$ UIB

To show that evolutionary stability does not imply that there is a uniform

invasion barrier, take the game  $A(x, y) = 3xy - x^4 - y^4$  and the strategy  $P$  degenerate in 0.<sup>4</sup> Since  $u[Q, P] < 0 = u[P, P] \forall Q \neq P$ , there is always an  $\epsilon_Q$  such that  $u[P, (1 - \epsilon)P + \epsilon Q] > u[Q, (1 - \epsilon)P + \epsilon Q]$  for all  $\epsilon \in (0, \epsilon_Q)$ . But for  $Q_z$  degenerate in  $z$  with  $z \rightarrow 0$ , we find, after rewriting the invasion barrier equation, that

$$\frac{\epsilon_{Q_z}}{1 - \epsilon_{Q_z}} = \frac{u[P, P] - u[Q_z, P]}{u[Q_z, Q_z] - u[P, Q_z]} = \frac{0 - (-z^4)}{3z^2 - 2z^4 - (-z^4)} = \frac{z^4}{3z^2 - z^4} = \frac{z^2}{3 - z^2},$$

which tends to 0 if  $z$  tends to 0. Therefore  $P$ , an evolutionary stable strategy, fails to have a uniform invasion barrier.

We can show the following though:

**Proposition 4** *If  $P$  is an ESS,  $S$  is compact and  $\{Q_n\}_{n \in \mathbb{N}}$  is a sequence of strategies with an invasion barrier that tends to 0,  $Q_n$  converges weakly to  $P$ .*

**Proof.** Let  $\{Q_n\}_{n \in \mathbb{N}}$  be a sequence such that  $\epsilon_{Q_n} \downarrow 0$ . In order to apply the lemma of Helly-Bray<sup>5</sup> we will identify probability measures  $Q_n$  with distribution functions  $G_n$ . Since  $S$  is compact, the sequence  $\{Q_n\}_{n \in \mathbb{N}}$  is naturally tight. With Helly-Bray we now know that there is a right-continuous non-decreasing function  $G$  and a subsequence  $\{G_{n_i}\}_{i \in \mathbb{N}}$  such that  $\lim_{i \rightarrow \infty} G_{n_i}(x) = G(x)$  at every point of continuity of  $G$ . The tightness of the sequence guarantees that this  $G$  is a distribution function, which we therefore can identify with a probability measure  $Q$ . Now assume  $Q \neq P$ .

Since  $P$  is an ESS, we know by definition that for every  $Q_n$  either  $u[P, P] > u[Q_n, P]$  or  $u[P, P] = u[Q_n, P]$  and  $u[P, Q_n] > u[Q_n, Q_n]$ . Because in the latter case the invasion barrier is 1 as well as in the case of  $u[P, P] > u[Q_n, P]$  with  $u[P, Q_n] \geq u[Q_n, Q_n]$ , the fact that  $\epsilon_{Q_n}$  decreases to 0 implies that from a certain  $N$  onwards all the strategies  $Q_{n_i}, i > N$  fall into the category  $u[P, P] > u[Q_n, P]$  and  $u[P, Q_n] < u[Q_n, Q_n]$ .

The invasion barrier equation tells us that

$$\lim_{i \rightarrow \infty} \frac{u[P, P] - u[Q_{n_i}, P]}{u[Q_{n_i}, Q_{n_i}] - u[P, Q_{n_i}]} = \lim_{i \rightarrow \infty} \frac{\epsilon_{Q_{n_i}}}{1 - \epsilon_{Q_{n_i}}} = 0.$$

Weak convergence of  $Q_{n_i}$  means that  $u[Q_{n_i}, P] \rightarrow u[Q, P]$ ,  $u[P, Q_{n_i}] \rightarrow u[P, Q]$  and  $u[Q_{n_i}, Q_{n_i}] \rightarrow u[Q, Q]$ , since  $A$  is continuous and  $S$  compact. But then we have found that  $P$  is not an ESS after all since  $u[P, Q] \leq u[Q, Q]$  and  $u[P, P] = u[Q, P]$ , which follows directly from the fact that  $u[Q_{n_i}, Q_{n_i}] - u[P, Q_{n_i}]$  is bounded.

Hereby we have shown that  $P$  is the only limit point of  $\{Q_n\}_{n \in \mathbb{N}}$ , and therefore that  $Q_n$  converges weakly to  $P$ . ■

This result may not be too telling all by itself, but it can be used to show that the following, more meaningful proposition holds.

<sup>4</sup> This example is from Oechssler and Riedel (2001), slightly adapted.

<sup>5</sup> See for instance Williams (1991).

**Proposition 5** *Assume contests are pairwise. If  $P$  has a density and  $S$  is compact,  $P$  has an UIB if  $P$  is an ESS.*

**Proof.** Assume that  $P$  is an ESS and that it has a density. Since  $P$  is an ESS, every strategy  $Q$  has an invasion barrier  $\epsilon_Q > 0$  and it will do to prove that there is no sequence  $\{Q_n\}_{n \in \mathbb{N}}$  such that  $\lim_{n \rightarrow \infty} \epsilon_{Q_n} = 0$ . To do so, assume that there is such a sequence. Proposition 4 tells us that then  $Q_n$  converges weakly to  $P$ . Now first consider the possibility that after a finite number of  $Q_n$ 's the remaining tail has only carriers that are subsets of the carrier of  $P$ , i.e. there is an  $N$  such that  $C(Q_n) \subset C(P)$  for all  $n > N$ . In that case the invasion barrier of every such  $Q_n$  equals 1;  $u[P, P]$  equals  $u[Q_n, P]$  since  $P$  is a Nash equilibrium and because  $P$  is also an ESS,  $u[Q_{n_i}, Q_{n_i}]$  must be smaller than  $u[P, Q_{n_i}]$ . But then  $\lim_{n \rightarrow \infty} \epsilon_{Q_n} = 1$  while we just assumed  $\epsilon_{Q_n}$  to go to 0, and from this contradiction we can conclude that there must be a subsequence  $\{Q_{n_i}\}_{i \in \mathbb{N}}$  for which  $C(Q_{n_i})$  is not a subset of  $C(P) \forall i \in \mathbb{N}$ . Now look at the sequence

$$\frac{\epsilon_{Q_{n_i}}}{1 - \epsilon_{Q_{n_i}}} = \frac{u[P, P] - u[Q_{n_i}, P]}{u[Q_{n_i}, Q_{n_i}] - u[P, Q_{n_i}]}$$

for which we can assume without loss of generality that  $u[P, P] > u[Q_{n_i}, P]$ .<sup>6</sup> Splitting every  $Q$  in a part on the carrier of  $P$  and on its complement, we get  $Q_{n_i} = Q_{n_i}^1 + Q_{n_i}^2$  with  $C(Q_{n_i}^1) \subset C(P)$  and  $C(Q_{n_i}^2) \cap C(P) = \emptyset$  for every  $i$ . Every such part  $Q_{n_i}^j$  can be rescaled to make it a probability measure by defining  $\bar{Q}_{n_i}^j = Q_{n_i}^j(S)^{-1} \cdot Q_{n_i}^j$ . To keep the upcoming formula's at least a little legible, we will write  $\alpha_{i,j}$  for  $Q_{n_i}^j(S)$ . Now there is nothing left to do but algebra:

$$\begin{aligned} \frac{u[P, P] - u[Q_{n_i}, P]}{u[Q_{n_i}, Q_{n_i}] - u[P, Q_{n_i}]} &= \frac{u[P, P] - u[Q_{n_i}^1 + Q_{n_i}^2, P]}{u[Q_{n_i}^1 + Q_{n_i}^2, Q_{n_i}^1 + Q_{n_i}^2] - u[P, Q_{n_i}^1 + Q_{n_i}^2]} = \\ &= \frac{u[P, P] - u[Q_{n_i}^1, P] - u[Q_{n_i}^2, P]}{u[Q_{n_i}^1, Q_{n_i}^1] + u[Q_{n_i}^1, Q_{n_i}^2] + u[Q_{n_i}^2, Q_{n_i}^1] + u[Q_{n_i}^2, Q_{n_i}^2] - u[P, Q_{n_i}^1] - u[P, Q_{n_i}^2]} = \\ &= \frac{\alpha_{i,2}(u[P, P] - u[\bar{Q}_{n_i}^2, P])}{(\alpha_{i,1})^2 u[\bar{Q}_{n_i}^1, \bar{Q}_{n_i}^1] - \alpha_{i,1} u[P, \bar{Q}_{n_i}^1] + (\alpha_{i,2})^2 u[\bar{Q}_{n_i}^2, \bar{Q}_{n_i}^2] - \alpha_{i,2} u[P, \bar{Q}_{n_i}^2] + \alpha_{i,1} \alpha_{i,2} (u[\bar{Q}_{n_i}^1, \bar{Q}_{n_i}^2] + u[\bar{Q}_{n_i}^2, \bar{Q}_{n_i}^1])} \end{aligned}$$

Here we used the bilinearity of  $u$  and the fact that  $u[P, P] = u[\bar{Q}_{n_i}^1, P]$ . Since  $Q_n$  converges weakly to  $P$ , we know that the same is true for  $\bar{Q}_{n_i}^1$  and that  $\lim_{i \rightarrow \infty} Q_{n_i}^2(S) = 0$ . Using this, we can say that the limit for  $i \rightarrow \infty$  of this expression equals

$$\lim_{i \rightarrow \infty} \frac{\alpha_{i,2}(u[P, P] - u[\bar{Q}_{n_i}^2, P])}{(\alpha_{i,1})^2 u[P, P] - \alpha_{i,1} u[P, P] + (\alpha_{i,2})^2 u[\bar{Q}_{n_i}^2, \bar{Q}_{n_i}^2] - \alpha_{i,2} u[P, \bar{Q}_{n_i}^2] + \alpha_{i,1} \alpha_{i,2} (u[P, \bar{Q}_{n_i}^2] + u[\bar{Q}_{n_i}^2, P])} =$$

---

<sup>6</sup> We could remove those  $Q$ 's for which this is not the case with the same procedure as we just used to thin out this sequence. This leaves us an infinite subsequence by the exact same argument; if not,  $\lim_{i \rightarrow \infty} \epsilon_{Q_i} = 1 \neq 0$ .

$$\lim_{i \rightarrow \infty} \frac{\alpha_{i,2} (u[P, P] - u[\bar{Q}_{n_i}^2, P])}{-\alpha_{i,1} \alpha_{i,2} u[P, P] + (\alpha_{i,2})^2 u[\bar{Q}_{n_i}^2, \bar{Q}_{n_i}^2] - (\alpha_{i,2})^2 u[P, \bar{Q}_{n_i}^2] + \alpha_{i,1} \alpha_{i,2} u[\bar{Q}_{n_i}^2, P]} =$$

$$\lim_{i \rightarrow \infty} \frac{(u[P, P] - u[\bar{Q}_{n_i}^2, P])}{-\alpha_{i,1} (u[P, P] - u[\bar{Q}_{n_i}^2, P]) + \alpha_{i,2} (u[\bar{Q}_{n_i}^2, \bar{Q}_{n_i}^2] - u[P, \bar{Q}_{n_i}^2])}$$

Since we assumed this limit to be 0 and because  $\lim_{i \rightarrow \infty} \alpha_{i,1} = \lim_{i \rightarrow \infty} Q_{n_i}^1(S) = 1$ , we obtain that<sup>7</sup>

$$\lim_{i \rightarrow \infty} \frac{(u[P, P] - u[\bar{Q}_{n_i}^2, P])}{\alpha_{i,2} (u[\bar{Q}_{n_i}^2, \bar{Q}_{n_i}^2] - u[P, \bar{Q}_{n_i}^2])} = 0$$

and therefore

$$\lim_{i \rightarrow \infty} \frac{(u[P, P] - u[\bar{Q}_{n_i}^2, P])}{(u[\bar{Q}_{n_i}^2, \bar{Q}_{n_i}^2] - u[P, \bar{Q}_{n_i}^2])} = 0$$

This though would imply that  $\lim_{i \rightarrow \infty} \epsilon_{\bar{Q}_{n_i}^2} = 0$ , but then  $d_{pr}(\bar{Q}_{n_i}^2, P)$  would go to 0 for  $i \rightarrow \infty$ , by proposition 4, which is not the case since  $P$  has a density and  $C(Q_{n_i}^2) \cap C(P) = \emptyset$ . Hence, the assumption that there is a sequence  $\{Q_n\}_{n \in \mathbb{N}}$  such that  $\lim_{n \rightarrow \infty} \epsilon_{Q_n} = 0$  leads to a contradiction when  $P$  has a density and  $S$  is compact. ■

## 2 UIB $\Rightarrow$ ESS.

The other direction is much easier; it is obvious that a strategy with a uniform invasion barrier is an ESS.

## 3 UIB $\nRightarrow$ LS<sub>kl</sub>

The following game is not very elegant, but it serves as an example of a game with a strategy that has a uniform invasion barrier without being locally superior in the Kullback-Leibler relative-entropy measure. Let  $S$  be the interval  $[-\pi, 2\pi]$  and take the sequence  $\{a_n\}_{n \in \mathbb{N}}$  with  $a_n = (1 - \frac{1}{2^n})\pi$ . Define the game as

$$A = \begin{cases} \left| \sin \frac{1}{2} (x - y) \right| & x \in [-\pi, \pi] & y \in [-\pi, \pi] \\ \frac{3(\pi - a_n) + 2 \cos \frac{1}{2} a_n}{(\pi + a_n)(a_n - a_{n-1})} & x \in [\pi + a_{n-1}, \pi + a_n] & y \in [-\pi, a_n] \\ 0 & \text{elsewhere} \end{cases}$$

Now take as an equilibrium strategy  $P$  the uniform distribution on  $[-\pi, \pi]$ , in other words,

---

<sup>7</sup> If  $a_n \neq 0 \forall n$ , and  $\lim_{n \rightarrow \infty} c_n = 1$  then  $\lim_{n \rightarrow \infty} \frac{a_n}{c_n a_n + b_n} = 0 \Leftrightarrow \lim_{n \rightarrow \infty} \frac{c_n a_n + b_n}{a_n} = \infty \Leftrightarrow \lim_{n \rightarrow \infty} c_n + \frac{b_n}{a_n} = \infty \Leftrightarrow \lim_{n \rightarrow \infty} \frac{b_n}{a_n} = \infty \Leftrightarrow \lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$

$$\frac{dP(x)}{d\lambda(x)} = \begin{cases} \frac{1}{2\pi} & x \in [-\pi, \pi] \\ 0 & \text{elsewhere} \end{cases}$$

where  $\lambda$  is the Lebesgue measure. Then we construct a sequence of signed measures  $\{H_n\}_{n \in \mathbb{N}}$  with  $H_n = H_n^+ - H_n^-$  which is defined by

$$\frac{dH_n^+(x)}{d\lambda(x)} = \begin{cases} \frac{1}{2\pi} & x \in [\pi + a_{n-1}, \pi + a_n] \\ 0 & \text{elsewhere} \end{cases} \quad \text{and} \quad \frac{dH_n^-(x)}{d\lambda(x)} = \begin{cases} \frac{1}{2\pi} & x \in [a_n, \pi] \\ 0 & \text{elsewhere} \end{cases}$$

This is a sequence of differences that we use to make a sequence  $\{Q_n\}_{n \in \mathbb{N}}$  of strategies, where  $Q_n = P + \frac{1}{2}H_n$ . Note that

$$\begin{aligned} \lim_{n \rightarrow \infty} d_{kl}(P, Q_n) &= \lim_{n \rightarrow \infty} \int \log \left( \frac{dP(x)}{dQ_n(x)} \right) dP(x) = \\ \lim_{n \rightarrow \infty} \left( \int_{-\pi}^{a_n} \log 1 dx + \int_{a_n}^{\pi} \log 2 dx \right) &= 0 \end{aligned}$$

With the obvious generalization of the function  $u$  to allow for the arguments to be measures that do not necessarily integrate to 1 as a shortcut in the algebra<sup>8</sup>, we can write

$$\begin{aligned} u[Q_n, Q_n] - u[P, Q_n] &= \\ u\left[P + \frac{1}{2}H_n, P + \frac{1}{2}H_n\right] - u\left[P, P + \frac{1}{2}H_n\right] &= \\ u[P, P] + \frac{1}{2}\{u[H_n, P] + u[P, H_n]\} + \frac{1}{4}u[H_n, H_n] - u[P, P] - \frac{1}{2}u[P, H_n] &= \\ \frac{1}{2}u[H_n, P] + \frac{1}{4}u[H_n, H_n] &= \\ \frac{1}{2}\{u[H_n^+, P] - u[H_n^-, P]\} + \frac{1}{4}\{u[H_n^+, H_n^+] - u[H_n^+, H_n^-] - u[H_n^-, H_n^+] + u[H_n^-, H_n^-]\} &= \\ \frac{1}{2}\left\{\frac{1}{4\pi^2}\left(3(\pi - a_n) + 2\cos\frac{1}{2}a_n\right) - \frac{1}{4\pi^2}\frac{1}{2}(4(\pi - a))\right\} + \frac{1}{4}\left\{\frac{1}{4\pi^2}\left(4(\pi - a) - 8\cos\frac{1}{2}a_n\right)\right\} &= \\ \frac{1}{8\pi^2}\left((\pi - a_n) - 2\cos\frac{1}{2}a_n\right) &> 0 \end{aligned}$$

This implies that the strategy  $P$  is not locally superior in the Kullback-Leibler relative-entropy measure. To show that the strategy  $P$  has a uniform invasion barrier it will do to look at the sequence  $\{\tilde{Q}_n\}_{n \in \mathbb{N}}$  with  $\tilde{Q}_n = P + H_n$ . For this sequence we find that

---


$$\begin{aligned} {}^8 u[H_n^-, P - H_n^-] &= \int_{-\pi}^{a_n} \int_{a_n}^{\pi} \left| \sin \frac{1}{2}(x - y) \right| \frac{1}{2\pi} dx \frac{1}{2\pi} dy = \frac{2}{\pi^2} \cos \frac{1}{2}a_n \\ u[H_n^-, H_n^-] &= \int_{a_n}^{\pi} \int_{a_n}^{\pi} \left| \sin \frac{1}{2}(x - y) \right| \frac{1}{2\pi} dx \frac{1}{2\pi} dy = \frac{1}{\pi^2} ((\pi - a_n) - 2\cos \frac{1}{2}a_n) \\ u[H_n^+, P] &= u[H_n^+, P - H_n^-] = \int_{-\pi}^{a_n} \int_{\pi + a_{n-1}}^{\pi + a_n} \frac{3(\pi - a_n) + 2\cos \frac{1}{2}a_n}{(\pi + a_n)(a_n - a_{n-1})} \frac{1}{2\pi} dx \frac{1}{2\pi} dy = \\ \frac{1}{4\pi^2} (3(\pi - a_n) + 2\cos \frac{1}{2}a_n) & \end{aligned}$$

$$\frac{\epsilon_{\tilde{Q}_n}}{1-\epsilon_{\tilde{Q}_n}} = \frac{u[P,P]-u[\tilde{Q}_n,P]}{u[\tilde{Q}_n,Q_n]-u[P,Q_n]} = \frac{-u[H_n,P]}{u[H_n,P+H_n]} =$$

$$\frac{-\left((3(\pi-a_n)+2\cos\frac{1}{2}a_n)-4(\pi-a_n)\right)}{(3(\pi-a_n)+2\cos\frac{1}{2}a_n)-8\cos\frac{1}{2}a_n} = \frac{(\pi-a_n)-2\cos\frac{1}{2}a_n}{3(\pi-a_n)-6\cos\frac{1}{2}a_n} = \frac{1}{3}$$

This implies that all the strategies  $\tilde{Q}_n$  have an invasion barrier of  $\frac{1}{4}$ . It can easily be checked that this is also a lower bound on the invasion barriers of all other mutant strategies, which means that we have found a uniform invasion barrier.<sup>9</sup>

There are two kinds of assumptions we can make on the forms of strategies under which equivalence can be restored. Please note that both propositions do actually even a bit more than merely repair the third arrow from figure 1; the first goes from evolutionary stability directly to local superiority in the Kullback-Leibler relative-entropy measure (1 and 3 in one step) and the second gives a condition under which the existence of a uniform invasion barrier implies local superiority in the variational distance (3 and 7).

**Proposition 6** *Assume that the carrier of  $P$  equals the action space  $S$ . Then  $P$  is locally superior in the Prohorov metric if  $P$  is an ESS*

**Proof.** This follows directly from the second definition of evolutionary stability that states that  $P$  is evolutionary stable if both  $u[P,P] \geq u[Q,P] \ \forall Q$  and  $u[P,P] = u[Q,P] \Rightarrow u[P,Q] > u[Q,Q] \ \forall Q \neq P$  hold. If the carrier of  $P$  equals  $S$  then  $u[P,P]$  must equal  $u[Q,P]$  and then the second part of the definition tells us that  $P$  must even be globally superior, and therefore that  $P$  will also be locally superior in the Prohorov metric. ■

**Proposition 7** *Assume that the carrier of  $P$  is finite. Then  $P$  is locally superior in the variational distance if  $P$  has a uniform invasion barrier*

**Proof.** The proof can be found in Oechsler and Riedel (2001). Note that they refer to a strategy as uninvadable when I say it has a uniform invasion barrier and that their strong uninvadability equals what I call local superiority in the variational distance. ■

#### 4 $\text{LS}_{kl} \Rightarrow \text{UIB}$

This implication is relatively easy and basically not very different from the finite case:

---

<sup>9</sup>In the game  $A$  we could replace  $3(\pi - a_n) + 2\cos\frac{1}{2}a_n$  by anything between  $4(\pi - a_n)$  and  $8\cos\frac{1}{2}a_n$ . We would only have to adapt the  $Q_n$ 's accordingly; if we take  $(1-\beta)*4(\pi - a_n) + \beta*8\cos\frac{1}{2}a_n$ ,  $\beta \in (0,1)$ , then we have to choose an  $\alpha$  that exceeds  $\beta$  in  $Q_n = P + \alpha H_n$ .

**Proposition 8** *A strategy  $P$  which is locally superior in the Kullback-Leibler relative-entropy measure has a uniform invasion barrier.*

**Proof.** First observe that

$$\begin{aligned} d_{kl}(P, (1-\epsilon)P + \epsilon Q) &= \int_{C(P)} \log \left( \frac{dP(x)}{d[(1-\epsilon)P + \epsilon Q](x)} \right) dP(x) < \\ \int_{C(P)} \log \left( \frac{dP(x)}{d[(1-\epsilon)P](x)} \right) dP(x) &= \int_{C(P)} \log \frac{1}{(1-\epsilon)} dP(x) = \log \frac{1}{(1-\epsilon)} \end{aligned}$$

$P$  being locally superior in the Kullback-Leibler relative-entropy measure means that there is a vicinity  $V_{kl}(P, \delta)$  of  $P$  such that for all  $R \in V_{kl}(P, \delta)$

$$u[P, R] > u[R, R]$$

Now take  $\bar{\epsilon} = 1 - e^{-\delta}$ . Using our first observation, we find that for all  $\epsilon < \bar{\epsilon}$  and all  $Q \in \mathcal{P}[S, \mathcal{B}]$  the following holds:  $d_{kl}(P, (1-\epsilon)P + \epsilon Q) < \delta$ . The local superiority then tells us that

$$\begin{aligned} u[P, (1-\epsilon)P + \epsilon Q] - u[(1-\epsilon)P + \epsilon Q, (1-\epsilon)P + \epsilon Q] &> 0 \Leftrightarrow \\ u[\epsilon P, (1-\epsilon)P + \epsilon Q] - u[\epsilon Q, (1-\epsilon)P + \epsilon Q] &> 0 \Leftrightarrow \\ u[P, (1-\epsilon)P + \epsilon Q] - u[Q, (1-\epsilon)P + \epsilon Q] &> 0 \end{aligned}$$

But then we have shown that  $\bar{\epsilon}$  is an uniform invasion barrier. ■

Now we will turn to how the three types of local superiority relate to each other.

## 5 $\mathbf{LS}_{vd} \not\Rightarrow \mathbf{LS}_{pr}$

To show that local superiority in the variational distance does not imply local superiority in the Prohorov-metric, take as a counterexample  $A(x, y) = 3xy - x^2 - y^2$  and  $P$  degenerate in 0.<sup>10</sup> Now take a look at mutant strategies of the form  $Q = (1-\epsilon)P + \epsilon H$ , where  $H$  is a probability measure which places no mass at 0. For any such mutant  $Q$ , we look at

$$\begin{aligned} u[P, Q] - u[Q, Q] &= \\ u[P, (1-\epsilon)P + \epsilon H] - u[(1-\epsilon)P + \epsilon H, (1-\epsilon)P + \epsilon H] &= \\ \epsilon^2 u[P, H] - \epsilon(1-\epsilon)u[H, P] - \epsilon^2 u[H, H] \end{aligned}$$

which, by the double symmetry of  $A$  equals

$$\epsilon(2\epsilon - 1)u[P, H] - \epsilon^2 u[H, H]$$

---

<sup>10</sup>This example is from Oechssler and Riedel (2000), slightly adapted.



Since mixing only relatively deteriorates the payoff of the mutant, it will do to look at degenerate  $H$ , where  $H_z$  will be a probability measure degenerate in  $z$ .

$$\epsilon(2\epsilon - 1)u[P, H_z] - \epsilon^2 u[H_z, H_z] = (\epsilon - 3\epsilon^2)z^2$$

which is larger than 0 for  $\epsilon < \frac{1}{3}$ .  $P$  is therefore locally superior in the variational distance. Yet,  $u[Q_z, Q_z] = z^2 > -z^2 = u[P, Q_z]$ , which shows that  $P$  is not locally superior in the Prohorov-metric, since for any  $\epsilon$  there is a  $Q_z$  for which  $d_{pr}(Q_z, P) < \epsilon$ .

## 6 $\mathbf{LS}_{pr} \Rightarrow \mathbf{LS}_{vd}$

Since  $V_{vd}(P, 2\delta) \subset V_{pr}(P, \delta)$ , a vicinity in the Prohorov-metric contains one in the variational distance, so it is obvious that local superiority in the Prohorov metric implies local superiority in the variational distance.

## 7 $\mathbf{LS}_{kl} \not\Rightarrow \mathbf{LS}_{vd}$

For a counterexample we go back to the one we used at arrow **3** and modify it slightly. We change the game into

$$A = \begin{cases} \left| \sin \frac{1}{2}(x - y) \right| & x \in [-\pi, \pi] & y \in [-\pi, \pi] \\ \frac{8 \cos \frac{1}{2}a_n}{(\pi + a_n)(a_n - a_{n-1})} & x \in [\pi + a_{n-1}, \pi + a_n] & y \in [-\pi, a_n] \\ 0 & \text{elsewhere} \end{cases}$$

If we look at the same equilibrium strategy  $P$ , that is  $P$  uniform on  $[-\pi, \pi]$ , and at mutant strategies  $Q_n(\alpha) = P + \alpha H_n$ , with the  $H_n$ 's as before, we see that

$$\begin{aligned} & u[Q_n(\alpha), Q_n(\alpha)] - u[P, Q_n(\alpha)] = \\ & \alpha u[H, P] + \alpha^2 u[H, H] = \\ & \alpha(u[H^+, P - H^-] + u[H^-, P]) + \alpha^2(u[H^-, H^-]) = \\ & \alpha\left(\frac{2}{\pi^2} \cos \frac{1}{2}a_n - \frac{1}{\pi^2}(\pi - a_n)\right) + \alpha^2\left(\frac{1}{\pi^2}(\pi - a_n) - \frac{2}{\pi^2} \cos \frac{1}{2}a_n\right) \end{aligned}$$

and this is smaller than 0 if  $\alpha$  is smaller than 1. But for  $\alpha = 1$ , which makes the only sequence  $\{Q_n(\alpha)\}_{\pi \in \mathbb{N}}$  that goes to 0 in the variational distance but not in the Kullback-Leibler relative entropy measure, this expression equals 0. Therefore  $P$  is not locally superior in the variational distance, although it is locally superior in the Kullback-Leibler relative entropy measure.

## 8 $\mathbf{LS}_{vd} \Rightarrow \mathbf{LS}_{kl}$

As in 6,  $V_{kl}(P, \delta^2) \subset V_{vd}(P, \delta)$  means that local superiority in the variational distance implies local superiority in the Kullback-Leibler relative entropy measure.

Before making the final steps, we will have to consider in what way definitions of asymptotic stability make sense. As with local superiority, the choice of a topology makes all the difference, but now not all topologies lead to sensible definitions. Starting with the first candidate, the Prohorov-metric, it is clear that for probability measures with isolated atoms, any neighbourhood contains points from which return is impossible; take as an example  $P$  degenerate in 0. Any neighbourhood does not only contain points of possible return, like a  $1 - \epsilon$  point mass at 0 and an  $\epsilon$  mass elsewhere, but also points that exclude ever getting back, like a probability measure degenerate in  $\epsilon > 0$ . Therefore asymptotic stability in the Prohorov-metric doesn't make all that much sense for this  $P$ . Asymptotic stability in the variational distance does make sense for this degenerate  $P$ , but not in general, as is indicated by another example. Take  $P$  uniform on  $[0, 1]$ . Again, there are points closeby from where a population could come back, but any neighbourhood also contains for instance probability measures that are uniform on  $[y, 1]$ ,  $y \in (0, 1)$ , from which there is no return.

This leaves us with the topology induced by the Kullback-Leibler relative entropy measure as the only remaining candidate. This is the strongest topology, but we need not worry about that too much; the cases in which asymptotic stability in the variational distance could also make sense are exactly those in which variational distance vicinities and Kullback-Leibler vicinities are basically equivalent.<sup>11</sup>

## 9 $\mathbf{LS}_{kl} \Rightarrow \mathbf{AS}_{kl}$

For showing that local superiority implies asymptotic stability in the Kullback-Leibler relative entropy measure, it is convenient first to establish that if we start at  $t = 0$  in an out of equilibrium point  $Q(0)$ , and watch the replicator dynamics do what it does, that then the time derivative of the Kullback-Leibler relative entropy measure at  $Q(t)$  equals  $-f_P(Q) = -(u[P, Q] - u[Q, Q])$ , which will of course prove to be a useful result. Observe first that

$$H_P(Q) = \int_{C(P)} \log\left(\frac{dP(x)}{dQ(x)}\right) dP(x) = - \int_{C(P)} \log\left(\frac{dQ(x)}{dP(x)}\right) dP(x) \geq 0$$

and then that for  $Q$  with  $H_P(Q) < \infty$

$$\begin{aligned} \frac{d}{dt} H_P(Q(t)) &= \frac{d}{dt} \left[ - \int_{C(P)} \log\left(\frac{dQ(x,t)}{dP(x)}\right) dP(x) \right] \Bigg|_{t=0, Q(x,0)=Q(x)} = \\ &= - \int_{C(P)} \frac{1}{dP(x)} \frac{dP(x)}{dQ(x,t)} \frac{\partial}{\partial t} (dQ(x,t)) dP(x) \Bigg|_{t=0, Q(x,0)=Q(x)} = \\ &= - \int_{C(P)} \frac{dP(x)}{dQ(x,t)} dQ(x,t) \left( \int_S A(x,y) dQ(y,t) - \int_S \int_S A(y,z) dQ(z,t) dQ(y,t) \right) \Bigg|_{t=0, Q(x,0)=Q(x)} = \end{aligned}$$

---

<sup>11</sup> That is: for any vicinity of the one type, one can find a vicinity of the other that is included in it.

$$\begin{aligned}
& - \int_{C(P)} \frac{dP(x)}{dQ(x,t)} dQ(x,t) \left( \int_S A(x,y) dQ(y) - \int_S \int_S A(y,z) dQ(z) dQ(y) \right) = \\
& - \int_S \left( \int_S A(x,y) dQ(y) dy - \int_S \int_S A(y,z) dQ(z) dQ(y) \right) dP(x) = \\
& - \int_S \int_S A(x,y) dQ(y) dP(x) - \int_S \int_S A(y,z) dQ(z) dQ(y) = - (u[P, Q] - u[Q, Q])
\end{aligned}$$

Then we are set to prove that:

**Proposition 9** *If  $P$  is locally superior in the Kullback-Leibler entropy measure, then it is asymptotically stable in the replicator dynamics (in the same measure).*

**Proof.** The proof is the same as the proof of the similar theorem for the finite case and can be found on page 100 in Weibull (1995). ■

10  $\mathbf{AS}_{kl} \not\Rightarrow \mathbf{LS}_{kl}$

The counterexample for the finite case of course still serves as an example that this implication does not hold. For this distorted Rock Scissors Paper game, I again refer to Weibull (1995), page 102.

## 4 Shocks, mutations and vicinities

Being correct is one thing, making sense is another. In the preceeding sections we focussed completely on being correct proving or negating all kinds of theoretical implications, without really bothering to check whether the concepts under examination actually have a sensible interpretation. Therefore I suggest to reflect a little on what the theory aims to describe.

The reason why we are interested in those theorems is that they might help us explain why we see certain strategies prevail in a population that is playing a game. The composition of such a population might change over time and we could see this evolution as a stochastic process with two ingredients: mutation and selection. Following Weibull (1995) we chose an approach that disentangles these two elements. As a selection process, we opted for the replicator dynamics that describe for every possible starting point how the composition of the population evolves under the influence of the relative succes of the different strategies. Mutations on the other hand are not explicitly modeled, but the relatively small shifts are supposedly being brought together in what we call neighbourhoods for the finite action space and vicinities for the more general case. The central concept is *asymptotic stability* and a strategy is said to be asymptotically stable if there is a vicinity such that the replicator dynamics guarantee a pull back to this strategy after any shift to a point in this vicinity. This concept is obviously qualitative; a strategy is either stable or unstable and how stable a strategy is, is not a question that this construct can help us answer. In other words, if we know that there is a vicinity in which all shifts are neutralized by the selection

process we still do not know if this only works for a tiny vicinity and therefore only for tiny shifts, or if the replicator dynamics also work as a piece of elastic for giant leaps. The leading example of both Kandori, Mailath & Rob (1993) and Young (1993) shows that this lack of discrimination can be unsatisfying. Suppose we have a coordination game with one risk dominant strategy; take for instance the game

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

A population is then determined by the share  $x_1$  that plays the first strategy. Now both  $x_1 = 1$  and  $x_1 = 0$  are asymptotically stable, but since the replicator dynamics take all populations left from  $x_1 = \frac{2}{3}$  to the latter equilibrium, one is tempted to say that this one is more stable than the other; the maximum shock it can overcome is twice as big as for the first equilibrium. Taking different, but explicit mutation processes, both Kandori, Mailath & Rob and Young look at the limit case where mutation probabilities become arbitrarily small and they both come up with another qualitative concept that does make a distinction between the two equilibria from the example; Kandori, Mailath & Rob introduce *long run equilibria* and Young defines *stochastically stable states*. One of their main conclusion is that even though both equilibria are evolutionary stable strategies, the process of mutation and selection spends considerably more time near the risk dominant equilibrium and as mutation probabilities go to 0 the limit distribution even puts probability 1 on this state.

The example shows that in the discrete case we need a specification of the mutation process once we want to distinguish one asymptotically stable state from another. Now if we turn to a continuous action space we need to be precise about mutations before we even get to asymptotic stability. With a finite action space the absence of a specific description of how mutations arise does not hamper us defining asymptotic stability; there is only one candidate for the type of set that should contain the smaller shifts. But in the continuous case we have seen that there are different types of vicinities with different properties. A decision on what type of vicinity is the most appropriate can only be taken if we are a bit more precise about what mutations we consider likely.

The reason why we restrict ourselves to vicinities anyway is that whatever population state  $P$  we consider, we can always dream up shifts to population states far out from which there is no return - for instance a shift to another equilibrium or one that drives a set of actions into extinction that we cannot ignore. However, if these shocks are very unlikely, we want our concept of stability to disregard them. Therefore a vicinity would preferably be a slim set that nonetheless captures the relatively likely transitions and we would like to speak of a state  $Q$  as being close to  $P$  if and only if a shift from  $P$  to  $Q$  is relatively probable. In Sections 2 and 3 we encountered three types of vicinities. We also found out that one of them, the Kullback Leibler type, has a practical advantage over the other two since it is the only type for which we can actually prove anything with respect to asymptotic stability. Whether the only workable type of asymptotic stability is also informative about a stochastic process of mutation and selection remains to be seen though. In the remainder of this

section we will compare the three options and I will not only argue that none of them is perfect, but also that there is indeed some sense in using the Kullback Leibler cross entropy measure. In order to make these comparisons we will first concentrate on what kind shocks the different types of vicinities cover.

The Prohorov metric induces the most inclusive type of vicinity. In this metric, unlike the others, two degenerate probability measures are as close as the points in which they are degenerate. With good reason Oechssler and Riedel (2000b) argue that this can capture quite realistic shifts; if we assume a game with a pure equilibrium and a homogenous population that plays this equilibrium, one can imagine that a change in the environment changes the payoffs and the equilibrium strategy just a little. In the after-shock situation, the whole population is just a little out of equilibrium and a good question is then whether it will drift further away or move towards the new equilibrium. Now the bad news is that - as we saw in Section 3, just before arrow 9 - homogenous populations do not move at all in the replicator dynamics, so whatever likely transitions we have added to our vicinities by using the Prohorov metric instead of for instance the variational distance, these can never be included in useful theorems; no conditions whatsoever can guarantee asymptotic stability in the Prohorov metric unless they exclude those cases that make out the attraction of the metric. This immobility certainly is a reason to doubt the appropriateness of an approach that separates mutation and selection with this kind of shocks, for one can very well think of trajectories of the stochastic process that we aim to characterize that do not suffer from this inertia. Assume for instance that we start off in a completely mixed strategy with positive density on the whole action set and no atoms. Now even if the replicator dynamics would converge to a degenerate distribution, it would never get there in finite time. All points along the way would therefore still be distributions with positive densities on the whole action set, so if the game and therefore the equilibrium slightly changes, nothing would stop the process from starting to converge to the new equilibrium. Looking at the discrepancy between the concept of asymptotic stability in the Prohorov metric and what we expect from the actual process of mutation and selection we can conclude that for this type of shocks our setting leads to a dead end.

Then we are left with only two candidates: the variational distance and the Kullback Leibler cross entropy measure. If we confine ourselves to vicinities in the variational distance we get a problem similar to the one we just saw in the Prohorov metric. This problem does not occur anymore in population states with finite support for there one can find vicinities that exclude extinction of substantial parts with the variational distance. But for a distribution  $P$  with a density the example in Section 3 shows that however small a vicinity one chooses, as long as it is of the variational distance type, it will always contain states in which too large a subset of the action space is extinct to allow for a return to  $P$ . Let me rephrase that formally. For any population state  $P$  that has a density and for any  $\epsilon > 0$  one can always find a population state  $Q$  and a subset  $B$  of the action space such that  $P(B) > 0$ ,  $Q(B) = 0$  and  $d_{vd}(P, Q) = \epsilon$ . This is not the case for the Kullback Leibler cross-entropy measure. If  $B$  is a subset of the action space  $S$  and  $P(B) > 0$ , then all probability measures in which all types

in the set  $B$  are extinct are infinitely far away from  $B$  in the Kullback Leibler cross-entropy measure and therefore not contained in any vicinity of  $P$  of this type. One could therefore say that in the Kullback Leibler cross entropy measure any population is extinction-by-mutation-proof and it is exactly this property that allows us to define asymptotic stability in a meaningful way. Going from a variational distance type vicinity to a Kullback Leibler type we apparently cut away the extinction cases and we are to decide whether it indeed is right to ignore these transitions. In order to be able to answer that question we will give a finger exercise and two more elaborate models of mutation hoping that these models point out a winner. But before we go on we should be aware that no sensible model of mutations could ever serve as a straightforward motivation for vicinities of the Kullback Leibler type. The fact that the Kullback Leibler cross-entropy measure does not satisfy the triangle inequality disqualifies this measure because I think anyone would agree that the transition probability from  $P$  to  $Q$  should be bounded below by the probability of a route from  $P$  to  $Q$  through  $R$ .

For the finger exercise we will consider the simplest action space possible and the most obvious mutation process. Our action space consists of two actions and every member of the population has an equal probability of mutating. With only two possible actions and  $N$  players, a state is described by the number of players that play the first action. Assuming a fixed mutation probability  $p$ , transition from a state  $x_1$  into  $y_1$  occurs with probability

$$\sum_{n=0}^{\min\{x_1, y_2\}} \binom{x_1}{n} p^n (1-p)^{x_1-n} \binom{x_2}{x_2-y_2+n} p^{x_2-y_2+n} (1-p)^{y_2-n} \quad \text{if } y_1 \geq x_1$$

$$\sum_{n=0}^{\min\{x_2, y_1\}} \binom{x_2}{n} p^n (1-p)^{x_2-n} \binom{x_1}{x_1-y_1+n} p^{x_1-y_1+n} (1-p)^{y_1-n} \quad \text{if } y_1 \leq x_1$$

where  $x_2 = N - x_1$  and  $y_2 = N - y_1$

A method to make a perfect vicinity - slim but comprehensive - would be to choose a set such that the transition probability to any point inside it exceed the transition probability to any point outside this set. In this case that would make a vicinity of a point  $x_1$  an interval  $(a, b)$  such that the transition probability from  $x_1$  to any point inside this interval is larger than the transition probability from  $x_1$  to any point outside this interval.<sup>12</sup> Following this procedure for a fixed  $p$  and  $N$ , we get sets that are skewed towards the middle and the further  $x_1$  is from the center, the more skewed its vicinities are.<sup>13</sup> This is a bit like the Kullback Leibler type vicinities. On the other hand, if we let  $p$  go to zero and  $N$

---

<sup>12</sup>It would be especially nice if we could find a function  $f$  such that  $f(x_1, c) > f(x_1, d)$  if and only if the transition probability from  $x_1$  to  $c$  is smaller than the one from  $x_1$  to  $d$ .

<sup>13</sup>For  $x_1 \neq \frac{N}{2}$  there is a size  $\epsilon$  such that for vicinities smaller than  $\epsilon$ ,  $x_1$  is not contained in its own vicinity. This undesirable property is due to the autonomous drift towards the middle that this model implies.

to infinity, these vicinities converge to exactly those we get with the variational distance.

Even in this simple model we can see why extinction-by-mutation proofness can be a good thing. If a vicinity of  $x_1$  contains a state in which one action is extinct, then the selection process is merciless: return from this action is impossible. This is the case even though mutation away from extinction is the most probable transition in the mutation process; the probability of going from  $x_1 = 0$  to  $x_1 = 1$  is an upper bound on all transition probabilities. Therefore I would argue that if our concept of stability qualifies extinction as an absorbing state of the selection process, even though the transition by mutation into extinction is a lot less likely than the other way round, it might be better to exclude the extinction cases from our vicinities. In this simple setting we could do this for the variational distance by simply choosing a smaller vicinity of the same type, but with a density on a continuous action space this no longer solves the problem, as we saw. Here the Kullback Leibler cross-entropy measure helps out.<sup>14</sup>

Generalizing our finger exercise can render us a model of mutations. If we divide the action space in  $m$  parts and let  $m$  go to infinity, this can be a way to approach a continuous action space. Further assumptions are more or less obvious: we assume that every member of the population has a small but fixed probability  $\bar{p}$  of mutating and that, given that it mutates, all actions - including the original action itself - are equally likely to be drawn. The probability of a transition from  $x = [x_1, \dots, x_m]$  to  $y = [y_1, \dots, y_m]$  then becomes:

$$\sum_{\substack{z \leq x \\ z \leq y}} \left[ \prod_{i=1}^m \binom{x_i}{z_i} \bar{p}^{(x_i - z_i)} (1 - \bar{p})^{z_i} * \frac{(N - \sum_{j=1}^m z_j)!}{\prod_{j=1}^m (y_j - z_j)!} \left(\frac{1}{n}\right)^{N - \sum_{j=1}^m z_j} \right]$$

where  $z \leq x$  if  $z_i \leq x_i \forall i$  and  $N = \sum_{i=1}^m x_i$ . If we want to be consistent with the simple model we can replace  $\bar{p}$  by  $\frac{m}{m-1}p$  which makes  $p$  the probability of mutating where all *other* actions are equally likely to be drawn if a player mutates. Perfect vicinities can be made by comparing transition probabilities in the same way we did in our simple model. The mutations cause a drift towards the uniform distribution and if we want not to be troubled by such complications we can, for the sake of simplicity start off in a uniform distribution. Perfect vicinities turn out to be somewhat between the vicinities we get from the variational distance and the Kullback Leibler cross-entropy measure, depending on what  $\bar{p}$ ,  $m$  and  $N$  we choose.

---

<sup>14</sup>Of course extinction should be an absorbing state in some cases; if one strategy is to be a tiger and the other to be a rhino and the tigers become extinct, they will never ever return. But this example does not exactly fit our model of mutation; we can hardly assume tigers to mutate into rhinos and vice versa. Especially if we turn to the case where the two types of vicinity really differ - distributions with a density - a continuous action space suggests a mutation process that would keep the carrier of an equilibrium strategy from lasting extinction. For instance, one could think of an action space representing a player's height or the moment at which a flower opens or a preference parameter. In those cases one expects that a next generation's action is an imperfect copy of the action of a current generation.

Using  $\sum_{i=1}^n (-1)^{i-1} \binom{n}{i} = 1$  we can compute the probability that there is an action that becomes extinct:

$$\sum_{i=1}^m (-1)^{i-1} \binom{m}{i} \left(\frac{m-i}{m}\bar{p}\right)^{\frac{iN}{m}} \left(1 - \frac{i}{m}\bar{p}\right)^{\frac{(m-i)N}{m}}$$

For an  $m$  that is not too large compared to  $N$  this probability is small compared to the probability of mutating away from extinction, even if we correct for the shrinking share of distributions with extinction in the total set of possible distributions. This could be used as an argument that lasting extinction can be so improbable that we indeed justly disregard that possibility.

There are also other ways to look at how, in a model with a continuous action space, large populations are approximated. Since populations in real life will always be finite, however large they are, one can say that at a given moment in time there is only a finite number of action being played, so in fact almost all actions are extinct. Nonetheless, we can treat these actions as if they were drawn from a distribution and what is more, if we assume that offspring of a player playing action  $\mu$  plays an action that is for instance normally distributed with expectation  $\mu$  and variance  $\sigma^2$ , where  $\sigma^2$  is very small, the aggregate actions of a next generation really is a distribution again. Such a model would be rather hard to handle, but in this setting we need not have too much fear of the absorbing character of extinction since the distribution of the action played by a next generation always has a density over the whole action space.

This second model of mutations also solves the problem of extinction in a case where the generalized finger exercise does not help. Consider for instance the distribution  $P$  on  $[0, 1]$  that is given by its density  $f(x) = 2x$  and assume it is an equilibrium. One can imagine that in our first setting the nearer an action is to 0, the closer to 1 the probability of it becoming extinct is. Extinction by mutation would therefore even be a very likely event. In this second setting however the population state can very well float around the equilibrium, since the crumbling away that once in a while occurs does not necessarily make the distribution of a next generations actions put probabilities on subsets of the action space that are very much different from the one from which the current generation is drawn.

Returning to the beginning of this section, one could argue that stability of a dynamic process would at its best be a quantitative notion that tells us how unlikely it is to depart from an equilibrium. It is not wide of the mark to state that trying to catch it in a qualitative definition such as asymptotic stability will never be a hundred percent satisfactory, which is even a little more apparent with a continuous action space. On the other hand, after what can be seen as a shortcut, we do arrive at a concept that makes some sense and that certainly has a useful generality to it, since even if a quantitative approach would render numbers at all, these would depend on size of the population, mutation probabilities and the speed of the selection process. The only workable type



of asymptotic stability is constructed with the Kullback Leibler cross-entropy measure and with different plausible models of mutation it is quite right in ignoring the cases where mutation leads to irreversible extinction.

## 5 Adaptive dynamics

Another approach to tackle games with a continuous action space in an evolutionary setting is taken by Hofbauer and Sigmund (1998). Their most important assumption is that the population will be homogeneous at every point along the timepath. This homogeneous population moves in the direction of the most advantageous mutation (of course from the individual point of view). A formal definition of the adaptive dynamics is

$$\frac{dx(t)}{dt} = \left. \frac{\partial A(y, x)}{\partial y} \right|_{y=x}$$

where  $x$  and  $y$  are to be thought of as vectors,  $x(t)$  representing the strategy the population plays at time  $t$ .  $A$  then needs to be differentiable and  $dx(t)/dt$  points in the direction of a vector  $h$  with length one that maximizes the expression  $A(x + \epsilon h, x) - A(x, x)$  for  $\epsilon \rightarrow 0$ . Furthermore they allow for the possibility of using other metrics than the Euclid metric, reflecting that some mutations can be more likely than others.

Note that their model is to capture both mutation and selection, whereas the replicator dynamics only focusses on selection. Homogeneous populations therefore do not move at all in the replicator dynamics, whereas in the adaptive dynamics there's nothing but (moving) homogeneous populations. I think a weak point of the adaptive dynamics is that it consists of no more than a definition, without an a priori reason why this would be the correct dynamics nor a specification of the mutation and selection proces that would (in the limit) lead to such a dynamics. Doing so, it overlooks the possibility of heterogeneous equilibria. For example, if we take  $S = [-\pi, \pi]$  and  $A(x, y) = \left| \sin\left(\frac{1}{2}(x - y)\right) \right|$ , adaptive dynamics predicts no movement at all in any homogeneous population, even though in any homogeneous population every mutation would do better than the incumbent strategy.

## A Appendix

In Section 2 we have, for different functions  $d$ , defined collections  $\mathcal{V}(P) = \{V(P, \delta)\}_{\delta > 0} = \{Q \in \mathcal{P}[S, \mathcal{B}] \mid d(P, Q) < \delta\}_{\delta > 0}$  of all vicinities of  $P$  for this  $d$ . Here we will put these sets in their proper topological context. Possibly superfluous, we first give a definition of a topological space and one of a neighbourhood:

**Definition 10** *A topological space is a pair  $(X, \mathcal{T})$ , with  $X$  a set and  $\mathcal{T}$  a topology, that is, a set of subsets of  $X$  that satisfies the following axioms*

1.  $\emptyset, X \in \mathcal{T}$
2. If  $O_i \in \mathcal{T} \forall i \in I$ , then  $\bigcup_{i \in I} O_i \in \mathcal{T}$
3. If  $O_1, \dots, O_n \in \mathcal{T}$ , then  $\bigcap_{i=1}^n O_i \in \mathcal{T}$

**Definition 11**  $U$  is a neighbourhood of  $P \in X$  if there is a  $O \in \mathcal{T}$  such that  $x \in O \subset U$ .

These are not the only possible definitions though; while we start with a topology and define neighbourhoods after that, Sierpinski (1952) starts off with a definition of neighbourhoods that is a very general version of what we called vicinities and given certain restrictions on them he defines a topology to go with these neighbourhoods. His definition of a topology actually requires more than Definition 10 does, which is more or less standard nowadays; it for instance includes the first separation axiom. This difference in definitions could be the source of some confusion, as we will see.

A classic way of defining a particular topology on a set  $T$  is done by choosing a function  $d : T \times T \rightarrow \mathbb{R}_0^+$  and stating that  $O \in \mathcal{T}$  if and only if for all  $x \in O$  there is a  $\delta > 0$  such that  $V(x, \delta) = \{y \in T \mid d(x, y) < \delta\} \subseteq O$ . This  $d$  is normally taken to be a distance, but that is not a necessity to make this define a topology. Actually, any function will do, but not all functions induce useful topologies, which we are also about to find out. The set on which we would like to define topologies is the set  $\mathcal{P}[S, \mathcal{B}]$  of distributions on the action space  $S$ . For the function  $d$  we choose the Prohorov distance, the variational distance and the Kullback-Leibler cross-entropy measure and the topologies we get in this way are denoted by  $\mathcal{T}_{pr}$ ,  $\mathcal{T}_{vd}$  and  $\mathcal{T}_{kl}$  respectively. The first two are also known as the weak and the strong topology and one is tempted to christen the third as the very strong topology. The differences between the three topologies are a little more subtle than the differences between the three sets of vicinities, though the examples are basically the same.

To illustrate the difference between  $\mathcal{T}_{pr}$  and  $\mathcal{T}_{vd}$ , we can take  $S = [0, 1]$ ,  $P$  degenerate in 0 and  $V = \{Q \in \mathcal{P}[S, \mathcal{B}] \mid d_{vd}(P, Q) < 1\}$ . This is an element of  $\mathcal{T}_{vd}$  because the variational distance is a distance.<sup>15</sup> To see that it is not an element of  $\mathcal{T}_{pr}$ , take a look at the probability measures  $P_y$  degenerate in  $y$ . For all  $y \in (0, 1]$  we find that  $d_{vd}(P, P_y) = 2$  where  $d_{pr}(P, P_y) = y$ . Therefore none of these distributions  $P_y$  are elements of  $V$ , even though any open set in the topology induced by the Prohorov metric that contains  $P$  must also contain some of these  $P_y$ . More formally:  $d_{vd}(P, P_y) = 2$  implies that  $P_y \notin V \forall y \in (0, 1]$  and since for every  $\delta > 0$  we can simply choose a  $y \in (0, \delta)$ , there is no  $\delta > 0$  such that  $d_{pr}(P, Q) < \delta \Rightarrow Q \in V \forall Q \in \mathcal{P}[S, \mathcal{B}]$  and therefore  $V \notin \mathcal{T}_{pr}$ .

The difference between  $\mathcal{T}_{vd}$  and  $\mathcal{T}_{kl}$  is illustrated by taking  $S = [0, 1]$ ,  $P$  uniform on  $S$  and  $V = \{Q \in \mathcal{P}[S, \mathcal{B}] \mid d_{kl}(P, Q) < 1\}$ , which is an element of

<sup>15</sup>That is:  $d_{vd}(P, P) = 0, d_{vd}(P, Q) > 0$  for  $Q \neq P$ , and  $d_{vd}(P, Q) + d_{vd}(Q, R) \geq d_{vd}(P, R)$ . The triangle inequality guarantees that open balls or vicinities as we defined them are indeed open sets.

$\mathcal{T}_{kl}$ .<sup>16</sup> It is not an element of  $\mathcal{T}_{vd}$  though; if we take distributions  $P_y$  uniform on  $[y, 1]$ , we find that for all  $y \in (0, 1)$  the Kullback-Leibler entropy  $d_{kl}(P, P_y) = \infty$  while  $d_{vd}(P, P_y) = 2y$ . Obviously  $P_y \notin V$  for  $y \in (0, 1)$  and since for every  $\delta > 0$  there is a  $P_y$ ,  $y \in (0, 1]$  with  $d_{vd}(P, P_y) < \delta$  there is, again, no  $\delta > 0$  such that  $d_{vd}(P, Q) < \delta \Rightarrow Q \in V \ \forall Q \in \mathcal{P}[S, \mathcal{B}]$  and therefore  $V \notin \mathcal{T}_{vd}$ .

Together with the delimitations from section 2 this all boils down to  $\mathcal{T}_{pr} \subset \mathcal{T}_{vd} \subset \mathcal{T}_{kl}$ .

Now we finally come to the reason why we used sets of vicinities and not topologies, which will take some ugly notation. First take the set of all possible vicinities that come with a given function  $d$ . This can be written as

$$F = \bigcup_{P \in \mathcal{P}[S, \mathcal{B}]} \mathcal{V}(P) \quad \text{where} \quad \mathcal{V}(P) = \{V(P, \delta)\}_{\delta > 0} \quad \text{and} \\ V(P, \delta) = \{Q \in \mathcal{P}[S, \mathcal{B}] \mid d(P, Q) < \delta\}.$$

A collections of this type is sometimes called a Frechét ( $V$ ) space. Now we will compare this to the topology we get in the way we mentioned above:

$$\mathcal{T} = \{O \mid \forall Q \in O \exists \delta > 0 : V(Q, \delta) = \{R \in \mathcal{P}[S, \mathcal{B}] \mid d(Q, R) < \delta\} \subseteq O\}$$

If this function satisfies the triangle inequality, every element of  $F$  is also an element of  $\mathcal{T}$ ; for every  $P \in \mathcal{P}[S, \mathcal{B}]$  and every  $\delta > 0$ , any point  $Q$  in  $V(P, \delta) = \{Q \in \mathcal{P}[S, \mathcal{B}] \mid d(P, Q) < \delta\}$  has a vicinity of its own that is a subset of  $V(P, \delta)$ ; take  $V(Q, \delta - d(P, Q)) = \{R \in \mathcal{P}[S, \mathcal{B}] \mid d(Q, R) < \delta - d(P, Q)\}$ . For a function  $d$  that is not a proper distance, this is not guaranteed and indirectly Csiszár (1962, 1964) shows that for the Kullback-Leibler entropy, which indeed does not satisfy the triangle inequality, not all vicinities are elements of the topology  $\mathcal{T}_{kl}$ . His result is actually even a bit stronger; there are probability distributions  $P$  for which no vicinity  $V(P, \delta)$ ,  $\delta > 0$  is an element of  $\mathcal{T}$ . The counterexample comes in the form of a probability distribution  $P$  and sequences of distributions  $\{Q_m\}_{m \in \mathbb{N}}$  and  $\{R_{m,n}\}_{n > m, m \in \mathbb{N}}$  for which  $\lim_{m \rightarrow \infty} d_{kl}(P, Q_m) = 0$  and  $\lim_{n \rightarrow \infty} d_{kl}(Q_m, R_{m,n}) = 0 \ \forall m$  but  $d_{kl}(P, R_{m,n}) = \infty \ \forall m, n$ .

To be perfectly complete, we give the definition of a basis of a topology and a necessary and sufficient condition for a set to serve as a basis for a topology

**Definition 12** *Let  $(X, \mathcal{T})$  be a topological space. A collection  $\mathcal{B}$  is a basis for  $\mathcal{T}$  if*

1.  $\mathcal{B} \subseteq \mathcal{T}$
2. for every  $O \in \mathcal{T}$  there is a  $\mathcal{B}' \subseteq \mathcal{B}$  such that  $O = \bigcup \mathcal{B}'$

**Theorem 13** *Let  $X$  be a set and  $\mathcal{B}$  a collection of subsets of  $X$ . Then there is a unique topology  $\mathcal{T}$  on  $X$  such that  $\mathcal{B}$  is a basis for  $\mathcal{T}$  if and only if*

---

<sup>16</sup>The Kullback-Leibler cross-entropy measure is not a distance, but for this particular  $P$  this vicinity is an open set;

1.  $\bigcup \mathcal{B} = X$
2. for all  $B_1, B_2 \in \mathcal{B}$  and all  $x \in B_1 \cap B_2$  there exists a  $B \in \mathcal{B}$  such that  $x \in B \subseteq B_1 \cap B_2$

The way Csiszár (1962, 1964) actually phrases his result is that he claims that  $F = \bigcup_{P \in \mathcal{P}[S, \mathcal{B}]} \mathcal{V}(P)$  is not a basis for a topology. This is definitely true, since from his counterexample  $V(P, \delta)$ ,  $\delta > 0$  and  $V(Q_m, \epsilon)$ ,  $\epsilon > 0$  could feature as  $B_1$  and  $B_2$  and  $R_{m,n}$  for  $n$  large enough would be an  $x$  for which condition 2 is not met. But if one reads the proofs of Csiszár or the article by Bomze (1991) who refers to this result, one should be aware that Csiszár uses the definitions by Sierpinsky (1952). As a consequence of his reversed procedure of defining what topologies and neighbourhoods are and his stronger definition of a topology, all these concepts are different from ours and cannot be interchanged. But even though his theorems would not remain valid if one would replace what he calls bases, topologies and neighbourhoods by what we chose for definitions, his counterexample, as we saw, still shows where the shoe pinches / where the rub is.

## References

- [1] Bickel, P.J., C.A. Klaassen, Y. Ritov and J.A. Wellner, 1993, Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press.
- [2] Bomze, I. M., 1991, Cross entropy minimization in uninhabitable states of complex populations, Journal of Mathematical Biology 30, 73-87
- [3] Bomze, I.M. and B.M. Pötscher, 1989, Game theoretical foundations of evolutionary stability, Springer, Berlin Heidelberg New York
- [4] Csiszár, I., 1962, Informationstheoretische Konvergenzbegriffe im Raum der Wahrscheinlichkeitsverteilungen, Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei 7, 137-158.
- [5] Csiszár, I., 1964, Über topologische und metrische Eigenschaften der relativen Information der Ordnung  $\alpha$ . Transactions of the third Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, Prague 63-73.
- [6] Oechssler, J. and F. Riedel, 2000, On the Dynamic Foundation of Evolutionary Stability in Continuous Models, mimeo.
- [7] Oechssler, J. and F. Riedel, 2001, Evolutionary Dynamics on Infinite Strategy Spaces, forthcoming Economic Theory 17, 141-162.
- [8] Hofbauer, J. and K. Sigmund, 1998, Evolutionary Games and Population Dynamics, Cambridge University Press, Cambridge.

- [9] Reiss, R.-D., 1989, Approximate distributions of order statistics, Springer Verlag, Berlin.
- [10] Seymour, R., 2000, Dynamics for Infinite Dimensional Games, mimeo, University College London.
- [11] Sigmund, K., 1987, Game dynamics, mixed strategies and gradient systems, Theoretical Population Biology 32, 114-126.
- [12] Sierpinski, W., 1952, General Topology, University of Toronto Press, Toronto.
- [13] Weibull, J.W., 1996, Evolutionary Game Theory, MIT Press, Cambridge, MA.
- [14] Williams, D, 1991, Probability with Martingales, Cambridge University Press, Cambridge.